

UNIVERSIDAD DE ALICANTE

FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES



GRADO EN ECONOMÍA
CURSO ACADÉMICO 2017 – 2018

**EL IMPACTO DE LOS TITULARES DE LAS NOTICIAS EN EL MERCADO
BURSÁTIL ESTADOUNIDENSE**

MARCOS RUIZ ALONSO

PEDRO ALBARRÁN PÉREZ

FUNDAMENTOS DEL ANÁLISIS ECONÓMICO

Alicante, julio de 2018

Resumen

Las noticias económicas, políticas y sociales son una fuente de información que los inversores aprovechan para tomar sus decisiones en el mercado bursátil. Por este motivo, el objetivo del presente trabajo es analizar el impacto de las noticias sobre la evolución del precio de las acciones. Se ha empleado la técnica de análisis de sentimiento para extraer la información relevante de los titulares y se ha analizado su efecto sobre el precio de las acciones de empresas del índice S&P500 mediante el estimador de efectos fijos y MCO. Por último, se concluye que el contenido que aportan los titulares sí afecta a la evolución del precio de las cotizaciones.

Palabras clave: minería de texto, titulares de noticias, análisis de sentimiento, mercado bursátil y finanzas.

ÍNDICE

1.	INTRODUCCIÓN	4
2.	REVISIÓN DE LA LITERATURA	5
2.1.	Conceptos previos	5
2.2.	Trabajos previos	8
3.	DATOS	10
3.1.	Corpus de titulares de noticias de las empresas.....	10
3.2.	Corpus de titulares de noticias preclasificado	10
3.3.	Lexicón de sentimientos	11
3.4.	Datos comerciales de las empresas.....	12
4.	PRIMER ESTUDIO EMPÍRICO: ANÁLISIS DE SENTIMIENTO	12
4.1.	Metodología.....	12
4.2.	Resultados.....	16
4.3.	Base de datos definitiva.....	18
5.	SEGUNDO ESTUDIO EMPÍRICO: ANÁLISIS DEL EFECTO DE LAS NOTICIAS	19
5.1.	Metodología.....	19
5.2.	Resultados.....	21
6.	CONCLUSIONES	30
7.	BIBLIOGRAFÍA	32
8.	APÉNDICE.....	34

1. INTRODUCCIÓN

En las finanzas, los inversores utilizan la información estructurada y no estructurada para tomar mejores decisiones de compra y venta de productos financieros. Pero por la gran cantidad de datos no estructurados que proporcionan los mercados financieros, es necesario la ayuda de algoritmos para poder interpretar los textos de manera más adecuada y rápida.

El objetivo de este trabajo es estudiar el efecto medio que tienen las noticias sobre la evolución del precio de una acción individual cualquiera y el efecto individual sobre cada empresa. Concretamente se utiliza el sentimiento de los titulares de las noticias del índice bursátil Standard & Poor's 500, como variable resumen de la información del titular para evaluar el efecto sobre los precios. Mientras realizaba mis prácticas en una empresa de trading intra-diario me surgió el interés por analizar cómo las noticias pueden influir de en el precio de las acciones: cuándo tienen efectos positivos y cuándo negativo, qué efecto tiene sobre el precio de apertura, cómo evoluciona a lo largo de la sesión, etc. Además, resultaba una forma de aplicar métodos y conocimientos adquiridos durante mi formación. Las herramientas computacionales y la técnica relativamente novedosa de análisis de sentimiento proporcionan una gran ventaja a los profesionales que la utilicen ya que es posible extraer la información relevante de los documentos en menor tiempo y poder observar sus efectos tanto en finanzas como en cualquier otro sector.

El trabajo se estructura de la siguiente manera: en el segundo apartado se hará una revisión de la literatura que incluirá definiciones y trabajos relacionados. En el tercer apartado se describirán las fuentes primarias y secundarias empleadas para el estudio. En la sección cuarta se explicará la metodología empleada para el análisis de sentimiento y se presentarán los resultados de este primer estudio. En la sección quinta, se explicará la metodología empleada para la estimación del impacto de los titulares sobre los precios de cotización y se presentarán los resultados de este segundo estudio. Y para concluir, en el sexto capítulo se comentarán las conclusiones que se han extraído del estudio, en la séptima y octava se incluirá la bibliografía utilizada y un apéndice.

2. REVISIÓN DE LA LITERATURA

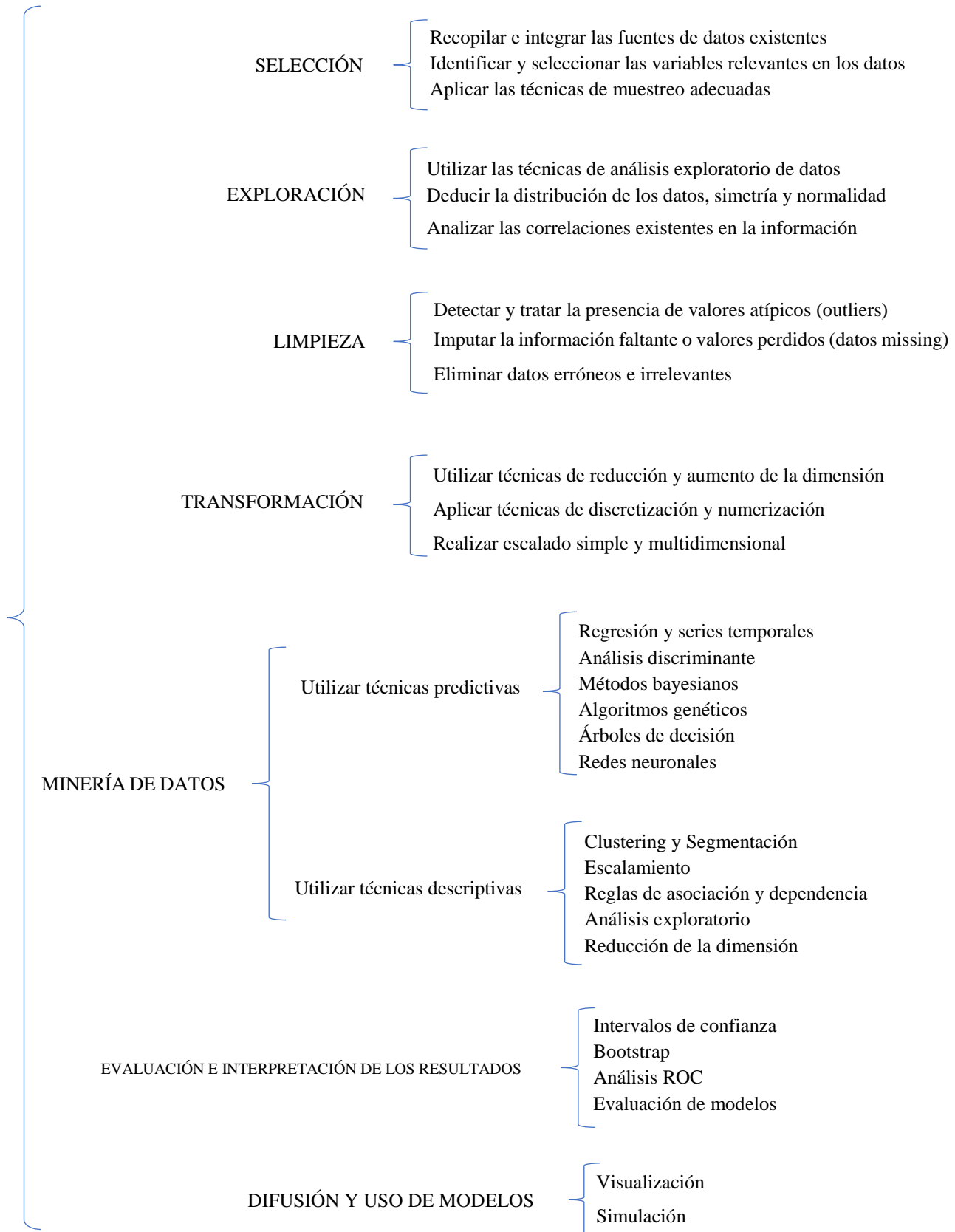
En este apartado se describen los conceptos en los que se basa el trabajo. También se describen trabajos que tienen relación con la minería de texto y el análisis de sentimiento, y que han inspirado el estudio, así como la metodología para obtener los resultados.

2.1. Conceptos previos

Hoy en día se generan cantidades masivas de datos a un ritmo vertiginoso desde múltiples fuentes y mediante complejos formatos, prácticamente de cualquier ámbito. Desde hace unos años se han definido a estas enormes cantidades de información como *Big data*. Hay varias definiciones para este término, una de ellas es la que proporciona la consultora Gartner, “Big data es un gran volumen, alta velocidad y / o gran variedad de activos de información que demandan formas rentables e innovadoras de procesamiento de la información que permiten un mejor conocimiento, toma de decisiones y automatización de procesos” (Gartner IT Glossary, 2018). Esta definición hace referencia a las 3 “V” (Volumen, Velocidad y Variedad) pero, adicionalmente se han propuesto nuevas “V” como Valor, Veracidad y Visualización entre otras.

El *Big Data* requiere de otros procedimientos para convertir a los datos en valor añadido, como por ejemplo la minería de datos. La disponibilidad de grandes volúmenes de información y el uso de herramientas informáticas han transformado el análisis de datos enfocándolo hacia determinadas técnicas especializadas, minería de datos o *Data Mining* (Pérez López and Santín González, 2008). Las diferentes técnicas y herramientas permiten extraer patrones, tendencias y regularidades para describir y comprender mejor los datos y poder predecir comportamientos futuros. La minería de datos es solo una parte del proceso denominado Descubrimiento de Conocimiento de Datos o *Knowledge Discovery from Data* (KDD), este proceso consta de diferentes fases. La primera fase es la preparación de los datos que consta de cuatro partes, selección, exploración, limpieza y transformación de los datos cuyo objetivo es eliminar erróneos e irrelevantes. La segunda fase es la minería de datos junto con la evaluación e interpretación de los resultados. Y la última fase es la difusión cuyo objetivo es que el modelo sea expresa de la forma más comprensible y clara. En la Figura 1 se puede observar en conjunto estas fases.

Figura 1: Fases del proceso de extracción del conocimiento (KDD)



Fuente: (Pérez and González, 2008)

La minería de datos ha ido modificándose durante los últimos años como consecuencia de los cambios tecnológicos. Una de las transformaciones más relevantes es la importancia que han cobrado los datos no estructurados (texto, páginas de internet, etc.). Surgen varias extensiones de la minería de datos, una de ellas es la minería de texto o *text mining* como consecuencia de la gran cantidad de datos no estructurados que se generan. Los datos no estructurados ocupan la mayor parte del espacio digital, el 80% del volumen corresponde a este tipo de datos (IBM Security, 2016).

La minería de texto es una técnica de la lingüística computacional y del procesamiento de textos cuyo objetivo es identificar y extraer nuevo conocimiento a partir de conjuntos de documentos (Valero Moreno, 2017). Existen varias aplicaciones para el *text mining*, dentro del procesado y extracción de textos se incluye, entre otros, la extracción de información o *feature extraction*, el análisis de sentimiento o minería de opiniones, la clasificación de documentos o *clustering* o la creación de resúmenes. (S. Baumer et al., 2017).

La extracción de información consiste en buscar en los textos nombres propios de personas, empresas, países, eventos o fechas, con el objetivo de encontrar una posible relación. Por ejemplo, de un documento se podrían obtener referencias a “Jerome Powell”, “presidente FED” y “tipos de interés”. Y encontrar relaciones entre estos como “Jerome Powell presidente de la FED” y “Jerome Powell sube los tipos de interés”. La clasificación se encarga de agrupar los documentos que se tiene según la similitud que exista entre ellos sin conocimiento previo de las agrupaciones y con el objetivo de facilitar la comprensión e identificar relaciones desconocidas. Por ejemplo, un resultado obtenido de la clasificación de noticias podría ser que las agrupara por sector. La elaboración de resúmenes se genera mediante la extracción de frases originales del documento, la selección de las frases se basa en la frecuencia de las palabras y de la posición que ocupan las frases y su objetivo es facilitar el análisis de grandes cantidades de documentos. La minería de opiniones es un campo de la investigación que se centra en clasificar las opiniones. Se puede decir que analizar el sentimiento del texto conlleva otorgar a cada documento una valoración relacionada con la carga emocional que intenta trasladar. En relación con los sentimientos se puede distinguir diferentes variables (Bravo-Marquez, Mendoza and Poblete, 2014):

- Polaridad: muestra si el texto contiene un sentimiento positivo o negativo. En otros análisis también se puede introducir una tercera clase para clasificar el texto neutro.

- Intensidad: otorga una valoración numérica relacionada con la intensidad del sentimiento. Se puede diferenciar entre una intensidad positiva y negativa.
- Emoción: valora el texto según las diferentes emociones, como la alegría o la tristeza entre otras.

El problema de clasificar el sentimiento de los textos ha sido abordado de diversas maneras. Existen principalmente dos formas distintas de enfrentarse a este problema (Liu, 2012): aplicando aprendizaje automático (Pang, Lee and Vaithyanathan, 2002) o aplicando un enfoque semántico (Turney, 2002).

El enfoque semántico utiliza diccionarios de palabras o lexicón con un determinado peso y/o categoría sentimental. En general, los sistemas pre-procesan el texto, dividen las palabras, eliminan las palabras de parada y lo normalizan lematizando las palabras. Y la forma de establecer si el sentimiento de una oración es positivo o negativo es mediante técnicas estadísticas y algoritmos que determinan la cantidad de palabras presentes del diccionario y la fuerza de su sentimiento. La ventaja de este enfoque es que se podría conseguir con relativa facilidad aumentar la precisión simplemente invirtiendo más tiempo en la construcción del lexicón. Sin embargo, crear un lexicón desde cero tiene un esfuerzo manual muy elevado.

El aprendizaje automático consiste en entrenar un clasificador usando un algoritmo de aprendizaje supervisado a partir de una colección de textos etiquetados previamente. Los sistemas más utilizados son los clasificadores basados en Support Vector Machine (SVM), Naive Bayes y K-Nearest Neighbor (KNN). La ventaja de este enfoque es que cuesta relativamente poco crear un analizador de sentimientos una vez tienes una colección de documentos etiquetados y se adaptan mejor al dominio planteado. Pero necesitan un reentrenamiento completo cuando quiere analizar un tema diferente.

2.2. Trabajos previos

La tarea que se pretende abordar es el análisis del sentimiento en el campo de los agregadores de noticias financieras y económicas, aunque existe una gran variedad de fuentes de datos a partir de los cuales se pueden emplear técnicas de análisis de sentimientos. Entre ellas cabe destacar los periódicos digitales, webs especializadas en críticas, blogs, redes sociales o sitios de microblogging como Twitter, Facebook o Reddit entre otras. Incluso algunas fuentes más específicas para el sector financiero como StockTwits o Stockwatch. Este tema ha recibido mucha atención dentro de la comunidad

NLP (Neuro-Linguistic Programing), por la importancia de la opinión pública y su repercusión en el ámbito político, económico y social. Hoy en día tanto empresas públicas como privadas se están centrando en aplicar las técnicas de análisis de sentimiento.

Una de las aplicaciones más interesantes y la cual ha motivado esta investigación es la relacionada con las empresas del sector financiero y el mercado de valores, como es el caso de la clasificación de opiniones negativas y positivas en Twitter para predecir el movimiento de los índices del mercado estadounidenses, como el Dow Jones (Zhang, 2013) y el S&P 500 (Sulaiman, 2017). También se ha demostrado que el poder predictivo de los microbloggings como Reddit es un poco mejor que el de los periódicos (Lubitz, 2018).

En el mundo financiero, el análisis de sentimiento tiene un alto impacto, ya que se ha demostrado que los sentimientos y las opiniones pueden afectar a la dinámica del mercado (Van De Kauter, Breesch and Hoste, 2015). En algunos casos, los sentimientos proceden de noticias que analizan factores macroeconómicos, información específica de la empresa o información política, ya que todos pueden ser relevantes para el mercado (Sinha, 2014). Las buenas noticias tienden a elevar los mercados y aumentar el optimismo de los inversores (Van De Kauter, Breesch and Hoste, 2015)

Dado el fuerte vínculo entre el sentimiento y el comportamiento del mercado, el análisis del sentimiento se convierte en un poderoso método para predecir la reacción del mercado. Aunque, el texto con una longitud reducida, como los mensajes de microblogs o los titulares de las noticias pueden ser bastante subjetivos y difíciles de analizar debido a los diferentes vocabularios utilizados. (Sinha, 2014). Por este motivo, la especificidad del tema que se quiere analizar afecta a todos los niveles. A nivel léxico es vital en el análisis de sentimiento porque las palabras positivas en un campo pueden ser negativas en otro y viceversa (Liu, 2016). Esto provoca que sea difícil de intercambiar clasificadores de sentimiento a través de diferentes campos y es importante utilizar herramientas específicas para cada tema.

3. DATOS

3.1. Corpus de titulares de noticias de las empresas

La base datos de los titulares de noticias de las empresas que cotizan en el mercado bursátil se han conseguido gracias a *The Shark Investment*. La empresa recoge sólo los titulares en inglés de las noticias mediante *Stockwatch*.

La plataforma se define como una fuente económica de suscripciones, operaciones y detalles del mercado en tiempo real para las empresas que cotizan en los principales mercados de valores de América del Norte, incluidos NYSE, AMEX, NASDAQ, OTCBB, OTC Markets, OPRA Options, Toronto, Montreal, CSE y CanDeal bonos canadienses. Stockwatch satisface las necesidades de inversores casuales, serios y profesionales con datos ilimitados en tiempo real del mercado, boletines de noticias de la compañía, diagramas dinámicos y herramientas analíticas.

Se almacenan todos los titulares de las noticias de cada día, antes de la apertura del mercado, es decir, desde primera hora de la mañana hasta las 15:30 (UTC+1), horario de apertura de New York Stock Exchange (NYSE). La base datos contiene titulares de empresas que cotizan tanto en NYSE como en Toronto Stock Exchange (TSE) desde el 2 de diciembre de 2016 hasta el 5 de abril de 2018, y se identifican mediante una “U” y una “C” respectivamente. Los días festivos tanto de Estados Unidos, Canadá como de España no se realiza la recolección de datos. En total la base de datos está formada por 154.462 titulares, de los cuales 119.610 (77,44%) pertenecen a empresas de NYSE y 34.852 (22,56%) pertenecen a empresas de TSE. Además, cada empresa se reconoce por su símbolo de cotización asociado.

Aparte de la fuente primaria de información, se buscan otro tipo de material léxico que es necesario para el análisis de sentimiento y la validación del modelo. Las fuentes secundarias se encuentran gracias a la revisión de la literatura y se describen a continuación. Además de descargar los datos comerciales de las empresas del índice bursátil.

3.2. Corpus de titulares de noticias preclasificado

En primer lugar, se recolectan y se agrupan dos corpus con titulares de noticias clasificados con sentimiento positivo y negativo. Se seleccionan estos corpus porque son similares a los titulares que se pretende clasificar ya que tienen una longitud y vocabularios similares.

El primero, se utilizó para la clasificación de la emoción de los titulares de noticias (Strapparava and Mihalcea, 2007) y la base de datos consiste en 1250 titulares de noticias extraídas de los principales periódicos como New York Times, CNN, la BBC y Google Noticias. Y el segundo, se utilizó para analizar el sentimiento en los titulares de las noticias de Reddit (Martin and Koufos, 2018) y la base de noticias que se ha podido recupera contiene 534 titulares. De 1784 titulares clasificados, 803 (45,01%) son positivos y 981 (54,99%) son negativos.

3.3. Lexicón de sentimientos

En segundo lugar, se buscan un lexicón para aplicarlo al análisis de sentimiento. Un lexicón o diccionario de palabras es un conjunto de palabras clasificadas según la polaridad, la intensidad o la emoción, entre otras. Fundamentalmente están formado por adjetivos que proporcionan la mayor información, también incluyen sustantivos, verbos y adverbios. Se encuentran y se unen 4 diccionarios con palabras clasificadas mediante el enfoque semántico.

El primero, fue compilado durante muchos años por los autores después de su primer trabajo (Hu and Liu, 2004) en el que se clasificaron las opiniones de los productos en la web. Contiene 6790 palabras de las cuales 4783 están clasificadas como negativas y 2007 como positivas (Liu, 2018).

El segundo, fue creado en el artículo (Chen and Skiena, 2014) cuyo objetivo era crear diferentes léxicos de alta calidad para los principales 136 idiomas porque los léxicos siguen siendo un recurso escaso para la mayoría de idiomas. Se elige el lexicón creado en inglés. Contiene un total de 4.376 palabras, de las cuales, 2.955 son negativas y 1.421 son positivas (Sites.google.com, 2018).

El tercero, fue diseñado con el objetivo de evaluar diferentes listas de palabras en el análisis de sentimiento de los microblogs. Hay dos versiones, AFINN-96 y AFINN-111 que contienen 1.468 y 2.477 palabras respectivamente y se han aplicado las dos (Nielsen, 2011).

Y el cuarto, se extrae de una web que proporciona recursos para el análisis del contenido de textos mediante un enfoque computacional (Mariapinto.es, 2018). Contiene 4.206 palabras, de las cuales, 1.915 son positivas y 2.291 son negativas.

3.4. Datos comerciales de las empresas

En tercer lugar, se descargan los símbolos de las empresas y sus respectivas cotizaciones pertenecientes al índice bursátil Standard & Poor's 500 (S&P500). Es un índice ponderado por capitalización bursátil de las 500 mayores empresas estadounidenses con acciones cotizadas en la Bolsa de Nueva York (NYSE) o NASDAQ. Es uno de los índices más seguidos por los inversores, y muchos los consideran una de las mejores representaciones del mercado bursátil de EE. UU. Aunque el índice se llama S&P "500", contiene 505 acciones porque incluye dos clases de acciones de 5 de sus compañías. Por ejemplo, la empresa matriz Google, Alphabet, tiene acciones de Clase A (GOOGL) y Clase C (GOOG) en el índice. En el trabajo se ha conseguido descargar todos los precios de las cotizaciones del índice excepto el de la empresa CBRE. Los datos comerciales contienen las siguientes variables:

- **Openn:** el primer precio de cotización tras la apertura del mercado bursátil.
- **Close:** el último precio de cotización antes del cierre del mercado bursátil.
- **High:** el precio de cotización más alto que alcanza la acciones durante la jornada bursátil.
- **Low:** el precio de cotización más bajo que alcanza la acciones durante la jornada bursátil.
- **Volumen:** el número total de transacciones que se han hecho durante la jornada bursátil.
- **Adjusted:** el precio resultante de aplicar al precio de cierre de una acción, el valor que implique el ejercicio de un derecho corporativo o patrimonial decretado por la emisora en cuestión.

El precio de apertura, cierre, máximo, mínimo y ajustado se miden en dólares estadounidenses.

4. PRIMER ESTUDIO EMPÍRICO: ANÁLISIS DE SENTIMIENTO

4.1. Metodología

Este trabajo se centra en el análisis de sentimiento porque con esta técnica se consigue extraer la información más relevante que se pretende transmitir en el documento analizado de una forma muy compacta y útil. Al igual que los titulares de las noticias son un resumen del contenido, el análisis de sentimiento es la única técnica que ofrece la posibilidad de crear una variable síntesis óptima que incluye la opinión que se desear

difundir. Concretamente el estudio se focaliza en la clasificación de la polaridad, es decir, en etiquetar el titular entre positivo, negativo o neutral y utilizar los negativos y positivos para explicar el efecto que tienen en la evolución de los precios de cotización de las empresas porque se espera que estos contengan información relevante.

Al ser una primera aproximación al análisis de sentimiento se utiliza el enfoque semántico por su relativa sencillez y buenos resultados, pero en lugar de crear un léxico, se ha utilizado diccionarios ya clasificados. Además, también se ha recopilado documentos pre-etiquetados para entrenar el algoritmo y evaluar los resultados. El algoritmo¹ utilizado consiste en contar el número de palabras positivas vs. negativas del documento que se pretende analizar y se asigna una polaridad, que puede ser neutral (0), negativa (<0) o positiva (>0) y estas dos últimas serán mayores o menores en función de la intensidad del sentimiento que exprese el texto analizado.

Antes de analizar el sentimiento de los titulares del índice bursátil, se evalúa la precisión del algoritmo de análisis de sentimiento mediante un corpus de titulares de muestra preclasificados utilizando las listas de palabras positivas y negativas. Se lematizan y se eliminan las palabras y titulares repetidos. Lematizar consiste en identificar las raíces o lemas de las palabras, al lematizar las palabras aumenta la exactitud del clasificador y elimina ruido del texto, es decir, las palabras que indican los mismo no son consideradas diferentes porque se eliminan las terminaciones morfológicas de las palabras (Im *et al.*, 2014). Se suman los diccionarios encontrados para aumentar el número de palabras y así poder ser obtener mayor precisión, en total hay 5.533 palabras negativas y 6.740 palabras positivas.

Se le aplica la función de análisis de sentimiento a los titulares preclasificados con el objetivo de comparar los resultados predichos con el verdadero valor de sentimiento. De esta forma se obtiene la exactitud con la que clasificarán los titulares de las noticias de las diferentes empresas del índice S&P500. Para la evaluación del algoritmo de clasificación se emplea una matriz de clasificación o confusión. La matriz de confusión es una herramienta estándar para la evaluación de modelos estadísticos (Visa *et al.*, 2011). Este instrumento, clasifica todos los casos del modelo en categorías y muestra en cada casilla el total de la suma de cada categoría. En este caso, las filas representan los valores reales y las columnas los valores predichos. Se utiliza esta técnica para evaluar los

¹ Breen, J. (2011). jeffreybreen (Jeffrey Breen). [online] GitHub. Available at: <https://github.com/jeffreybreen>

resultados de las predicciones porque facilita la comprensión y explicación de estas. En la tabla 1, se pueden observar los resultados del clasificador.

Tabla 1: Matriz de confusión

		Predicción	
		Negativo	Positivo
Real	Negativo	500 (VN)	108 (FN)
	Positivo	169 (FP)	296 (VP)

Fuente: Elaboración propia

Teniendo en cuenta que el interés de este trabajo se centra en 2 clases de titulares, los datos pueden ser agrupados en 4 conceptos:

- Verdadero Positivo (VP) = Cantidad casos positivos clasificados correctamente.
- Verdadero Negativos (VN) = Cantidad casos negativos clasificados correctamente.
- Falsos Positivos (FP) = Cantidad casos positivos clasificados incorrectamente.
- Falsos Negativos (FN) = Cantidad casos negativos clasificados incorrectamente.

Al tener en cuenta estos conceptos, se pueden calcular medidas para la evaluación de los resultados. Dos de las medidas más utilizadas para la evaluación de algoritmos de clasificación son las medidas *Precision*, *Recall* y *Accuracy* (Im et al., 2014).

- *Precision*: es el porcentaje de casos de una clase predicha que realmente son de esa clase. Se puede calcular tanto para la clase positiva como para la negativa utilizando las siguientes ecuaciones:

$$P_{pos} = \frac{VP}{VP + FP} \quad P_{neg} = \frac{VN}{VN + FN}$$

- *Recall*: es el porcentaje de casos de casos reales de la clase que se han clasificado como tal. También se puede calcular para las ambas clases de la siguiente forma:

$$R_{pos} = \frac{VP}{VP + FN} \quad R_{neg} = \frac{VN}{VN + FP}$$

- *Accuracy*: es el porcentaje del total de casos clasificados correctamente y se calcula así:

$$A = \frac{VP + VN}{VP + FP + VN + FN}$$

Si se analizan las variables de evaluación *precision* y *recall* se puede observar que existe un trade-off entre ambas variables, si se quisiera aumentar el *recall* se podrían recuperar, por ejemplo, se podría aumentar la cantidad de datos analizados, pero haciendo esto la *precision* puede disminuir considerablemente. Una medida que se ocupa de analizar la compensación entre ambas variables es el *F-score* y se define como la media armónica ponderada entre *precision* y *recall*, por tanto, esta variable tiene en cuenta tanto los FP como los FN.

$$F = \frac{(\beta + 1)PR}{\beta^2 P + R}$$

Valores de $\beta < 1$ acentúan la *precision*, mientras que valores $\beta > 1$ acentúan el *recall*. Se consideran que ambas medidas tienen el mismo peso, el estadístico F queda así:

$$F = \frac{2PR}{P + R}$$

La variable *F-score* generalmente es más útil que la variable *accuracy*, especialmente si tiene una distribución de clases desigual pero la variable *accuracy* es más conveniente si el valor de los FP y los FN son similares. Y si el valor de los FP y los FN es muy diferente es mejor mirar tanto las variables *precision* como *recall*.

Después de la evaluación de la precisión del modelo, se seleccionan todos los titulares de los que se dispone en la base de datos correspondientes a los símbolos pertenecientes al índice bursátil que son 14.248 y se les aplica el algoritmo de análisis de sentimiento para clasificarlos según la polaridad que transmitan. En la tabla 2 se puede observar que en la base de datos no se dispone de los mismos titulares para cada empresa.

Se crea una variable llamada “sent” en la que se recogen la polaridad de los titulares y como consecuencia de que no hay titulares de noticias cada día para todas las empresas, se decide tratar los valores faltantes sustituyéndolos por 0. Este tratamiento significa que la polaridad del titular es neutra y, por tanto, no tiene impacto en el precio de cotización del índice. Por este motivo se decide crear dos variables dummy para recoger sólo los titulares que se ha demostrado en otros trabajos que pueden influir en la evolución del precio de las cotizaciones.

Tabla 2: Estadísticos principales de titulares utilizados

Titulares por empresas	
Mínimo	2
1st Qu.	12
Mediana	22
Media	28
3rd Qu.	37
Máximo	199
N	14.248

Fuente: Elaboración propia

4.2. Resultados

En primer lugar, se muestran en la tabla 3 los resultados de las medidas de evaluación del algoritmo de clasificación de los titulares de prueba.

Tabla 3: Variables de evaluación

	Positivo	Negativo
Precision	63,66 %	82,24 %
Recall	73,27 %	74,74 %
F-score	68,13 %	78,31 %
Accuracy	74,18 %	

Fuente: Elaboración propia

De estos primeros resultados, se puede observar que los titulares con un sentimiento negativo es más probable que sean efectivamente titulares con polaridad negativa, a diferencia de los titulares positivos donde la probabilidad de que sean efectivamente titulares positivos disminuye. Observando los niveles de recall, se puede decir que la probabilidad de que se detecte la polaridad del titular mediante las palabras del diccionario es del 73,27% para los positivos y del 74,74% para los negativos. Al final estas dos variables son más fiables que *F-score* y *accuracy* porque en la base de datos de prueba, hay más titulares negativos que positivos y hay una diferencia de 61 titulares entre FP y FN. En comparación con otros trabajos, Im et al. (2014) consiguieron en su investigación utilizando el análisis de sentimiento con lexicón y sólo los titulares de las noticias un clasificador fiable entrono al 70%.

En segundo lugar, se muestran los estadísticos principales de la variable que recoge el sentimiento de los titulares, “sent”. La variable “sent” es una variable categórica que incluye tres clases de titulares, positivos, negativos con diferentes intensidades y neutrales. En la tabla 4 se muestra los detalles de la variable.

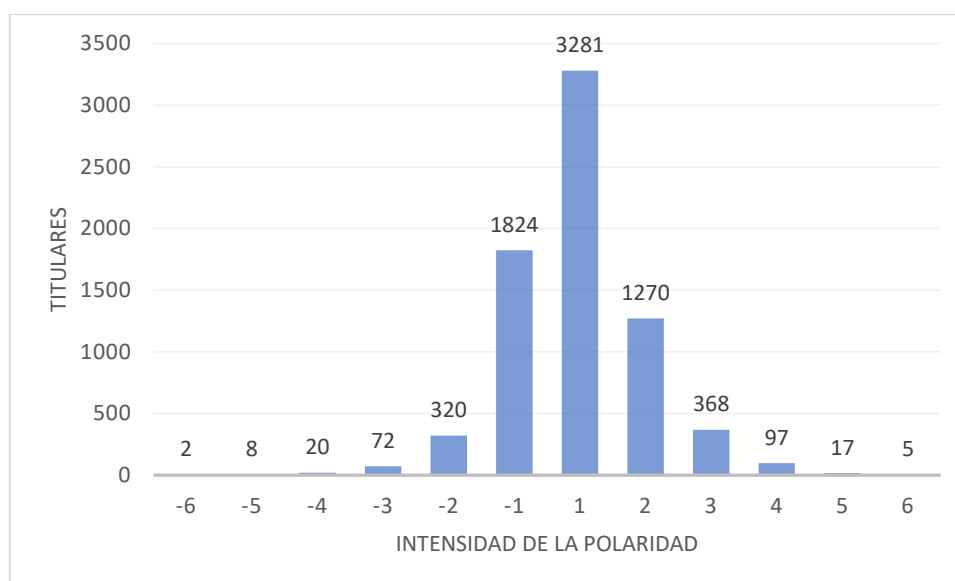
Tabla 4: Estadísticos principales de "sent"

	Variable “sent”
Número de observaciones (N)	168.987
Máximo	6
Mínimo	-6
Titulares positivos	5.038
Titulares negativos	2.246
Titulares neutrales	161.633

Fuente: Elaboración propia

Como se puede observar el sentimiento que más abunda es el neutral porque se ha incluido neutrales ficticiamente para sustituir los valores ausentes y por la dificultad de analizar ya que los titulares pueden ser muy subjetivos. Aun así, el algoritmo consigue clasificar el 51,12 % de los titulares que hay en la base de datos de las empresas del índice. Como los titulares de interés en este estudio son los positivo y negativos se crea la variable “pos” que recoger sólo los titulares positivos, vale 1 si el titular de la noticia es positivo y 0 en caso de ser neutral. La variable “neg” contiene sólo los titulares negativos, vale 1 si el titular de la noticia es negativo y 0 en caso de ser neutral. En el gráfico 1 se pueden observar las diferentes intensidades de polaridad de los titulares. Como se puede contemplar, los titulares más abundantes son los que tienen una polaridad levemente y medianamente positiva y negativa ya que hay muy pocos titulares catastróficos o sublimes.

Gráfico 1: Histograma de la intensidad de la polaridad



Fuente: Elaboración propia

4.3. Base de datos definitiva

A continuación, las tres variables relacionadas con el análisis de sentimiento que se han obtenido en el primer estudio empírico y que recogen la polaridad de los titulares de las noticias del índice bursátil se unen a los datos comerciales del precio de las cotizaciones de las empresas del S&P500. Se organizan con la estructura de datos de panel agrupados por símbolo y fecha. En la base datos definitiva se combina una dimensión temporal (serie de tiempo) y otra de corte transversal (individuos). La serie temporal se inicia el 12 de diciembre de 2016 y finaliza el 5 de abril de 2018 y los individuos son las empresas del índice bursátil. Además, se trata de datos de panel no balanceados porque hay fechas en las que el precio de cotizaciones de algunas empresas no se ha podido descargar. Este problema sucede con 8 empresas, de las cuales, 4 tienen datos para 122, 188 y 276 días y el resto para 335 días, es decir, solo les falta 1 ya que el total son 336 días. Como anteriormente se han observado los estadísticos principales de las variables de sentimientos, ahora se va a mostrar en la tabla 5 los estadísticos de los datos comerciales. Se puede apreciar una gran dispersión entre los diferentes precios ya que los rangos intercuartílicos son bastante grandes lo que indica que en el índice hay empresas muy heterogéneas y que sobre cada una de ella el efecto de los titulares de las noticias será diferente.

Tabla 5: Estadísticos principales de los datos comerciales

Variables	openn	high	low	close	volume	adjusted
Mínimo	0,11	0,11	0,10	0,11	0	-27,14
1st Qu.	45,99	46,41	45,55	46	1.082.200	45,19
Mediana	73,14	73,78	72,53	73,17	2.052.500	72,12
Media	98,58	99,45	97,69	98,59	4.073.786	101,53
3rd Qu.	116,75	117,65	115,80	116,76	4158200	115,30
Máximo	2.210,93	2.228,89	2174,07	2.206,09	312.556.800	36.148,86
N	168.987					

Fuente: Elaboración propia

5. SEGUNDO ESTUDIO EMPÍRICO: ANÁLISIS DEL EFECTO DE LAS NOTICIAS

5.1. Metodología

En este apartado, se va a explicar los modelos utilizados para estimar los efectos conjuntos e individuales de los titulares de las noticias sobre la evolución de los precios de las cotizaciones de las empresas del S&P500.

En primer lugar, para estimar el efecto conjunto que los titulares de noticias tienen sobre los distintos precios del mercado se emplea el modelo de efectos fijos por la forma de tratar los efectos individuales.

$$Y_{it} = \beta_1 X_{it} + \dots + \beta_k X_{kit} + u_{it}, \quad t = 1, 2, \dots, T. \quad (1)$$

$$Y_{it} = \beta_1 X_{it} + \dots + \beta_k X_{kit} + \alpha_i + \varepsilon_{it}, \quad t = 1, 2, \dots, T. \quad (2)$$

En el modelo, X_{1it}, \dots, X_{kit} son las variables explicativas, las que se pueden observar y en este trabajo son las variables dummy de sentimiento (pos y neg) y los distintos precios en de cotizaciones de las empresas del índice en logaritmos (l_openn, l_high, l_low, l_close, l_volumen y l_adjusted) e Y_{it} es la variable explicada, que también serán los distintos precios en logaritmos. Se utiliza el modelo log-log con el objetivo de estandarizar los datos, ya que no es lo mismo el efecto que puede provocar un titular en una empresa con precios elevados que una con precios bajos.

En este estudio, se va a intentar explicar el efecto que tienen los titulares de noticias sobre las siguientes variables dependientes: el precio de apertura, máximo, mínimo, cierre y el volumen. $u_{it} = \alpha_i + \varepsilon_{it}$ es el término de error del modelo y está compuesto por dos variables que recogen características inobservadas, la variable α_i es el efecto fijo o heterogeneidad no observada captura todos los factores individuales inobservables, constantes en el tiempo, que influyen en Y_{it} . Los efectos individuales pueden ser factores como por ejemplo tecnológicos, de calidad de un bien o la habilidad de la dirección. ε_{it} es el error idiosincrásico y captura factores inobservables que cambian en el tiempo, que son distintos a los que recoge los regresores y también influyen en Y_{it} .

El estimador de efectos fijos asume que el efecto individual α_i está correlacionado con las variables explicativas X_{1it}, \dots, X_{kit} . Este supuesto relaja la condición impuesta por el estimador de efectos aleatorios, tratando el efecto individual separadamente del término de error $Cov(X_{it}, \alpha_i) \neq 0$. Este estimador tiene la ventaja de que permite conocer los efectos individuales separadamente, lo que contribuye a entender mejor el modelo. Además, asumen que los regresores no están correlacionados con el error idiosincrásico ε_{it} , este supuesto $Cov(X_{it}, \varepsilon_{it}) = 0$ es fundamental para que el modelo estime de manera consistente los parámetros β_k .

Para explicar el funcionamiento de la estimación de efectos fijos, se utiliza la ecuación (2). Primero, para cada i , se promedia la ecuación en el tiempo y se obtiene:

$$\bar{Y}_i = \beta_1 \bar{X}_i + \dots + \beta_k \bar{X}_{ki} + \alpha_i + \bar{\varepsilon}_i \quad (3)$$

Donde $\bar{Y}_i = T^{-1} \sum_{t=1}^T Y_{it}$, y así sucesivamente. Como α_i permanece constante en el tiempo, aparece tanto en la ecuación (2) como en la (3). Si se resta la ecuación (3) de la ecuación (2) para cada t , se obtiene:

$$Y_{it} - \bar{Y}_i = \beta_1 (X_{it} - \bar{X}_i) + \dots + \beta_k (X_{kit} - \bar{X}_{ki}) + \varepsilon_{it} - \bar{\varepsilon}_i, \quad t = 1, 2, \dots, T,$$

o

$$\check{Y}_{it} = \beta_1 \check{X}_{it} + \dots + \beta_k \check{X}_{kit} + \check{\varepsilon}_{it}, \quad t = 1, 2, \dots, T. \quad (4)$$

Donde $\check{Y}_{it} = Y_{it} - \bar{Y}_i$ son los datos con el tiempo deducido sobre Y , y del mismo modo para \check{X}_{it} y $\check{\varepsilon}_{it}$. Lo importante respecto a la ecuación (4) es que el efecto inobservable, α_i , ha desaparecido (Wooldridge, 2010). Y, por tanto, los coeficientes de las variables dummy que contiene los titulares positivos (β_{pos}) y negativos (β_{pos}) captan los efectos de estos titulares en la variable dependiente.

En segundo lugar, para estimar el efecto individual que los titulares de noticias tienen sobre la evolución de los precios de su propia empresa se emplea el modelo de Mínimos Cuadrados Ordinarios (MCO).

$$Y_t = \beta_1 X_t + \dots + \beta_k X_{kt} + u_t, \quad t = 1, 2, \dots, T \quad (5)$$

En este modelo, las variables explicativas y explicadas son las mismas que antes, pero ahora se va a aplicar este modelo individualmente a cada empresa que forma el índice.

Esta segunda parte del estudio se aplica el modelo (5) a cada empresa y como antes, los coeficientes de las variables dummy que contiene los titulares positivos (β_{pos}) y negativos (β_{neg}) captan los efectos de estos titulares en la variable dependiente.

5.2. Resultados

En este apartado, se van a presentar los resultados de las estimaciones de los efectos conjuntos e individuales de los titulares de las noticias sobre el precio de las cotizaciones de las empresas del índice bursátil. El orden es el siguiente: primero el conjunto y a continuación el individual.

En primer lugar, para explicar el efecto de los titulares sobre la variable apertura se generan dos modelos. En el primero, se regresa la variable del precio de apertura frente a las dos variables binarias que recogen los sentimientos positivos y negativos dejando de lado los titulares neutrales, con el objetivo de diferenciar entre el efecto de los titulares de las noticias positivas y negativas que son las variables de mayor utilidad. En el segundo, se crea el mismo modelo, pero controlando por el primer retardo de las variables comerciales para explicar la evolución del precio de apertura con el objetivo de eliminar el sesgo por omisión de variable relevante. En la tabla 6 se pueden observar los resultados.

Tabla 6: Modelos precio apertura.

Variable dependiente	l_openn			l_openn		
	Coeficiente	D. Típica	P-valor	Coeficiente	D. Típica	P-valor
const	4,28694	0,00038	***	0,00538	0,00101	***
pos	0,01189	0,00222	***	0,00033	0,00012	***
neg	-0,00071	0,00329		-0,00014	0,00017	
l_high_1				0,01167	0,00389	***
l_low_1				0,02783	0,00364	***
l_close_1				0,97058	0,00347	***
l_volume_1				0,00006	0,00005	
l_adjusted_1				0,00074	0,00039	*
l_openn_1				-0,01213	0,00329	***
N	168.917			168.200		
R ²	0,960407			0,99988		

Fuente: Elaboración propia

Se puede observar en la tabla 6 que sólo la variable que recoge los titulares positivos afecta a la evolución del precio de apertura de las empresas del índice bursátil, por tanto, si el titular es positivo en lugar de ser neutral, el precio de apertura de las empresas del SPY aumentará en media 0,03%, permaneciendo constante el resto. Además, cabe destacar que el signo del coeficiente de la variable que recoge los titulares negativos es el esperado, aunque no sea significativo. Esto se puede deber a que hay pocos titulares negativos.

En segundo lugar, para explicar el efecto de los titulares sobre la variable cierre se generan otros dos modelos. En el primero, se regresa la variable del precio de cierre frente a las dos variables binarias que recogen los sentimientos positivos y negativos controlando por el primer retardo de las variables comerciales para explicar la evolución del precio de apertura con el objetivo de eliminar el sesgo por omisión de variable relevante. En el segundo, se crea el mismo modelo, pero incluyendo la variable del precio de apertura sin retardo como consecuencia de la recepción de información por parte de los inversores y con el objetivo de eliminar el sesgo por omisión de variable relevante. Es decir, los inversores conocen el precio de apertura el mismo día antes de que el mercado cierre. En la tabla 7 se pueden observar los resultados.

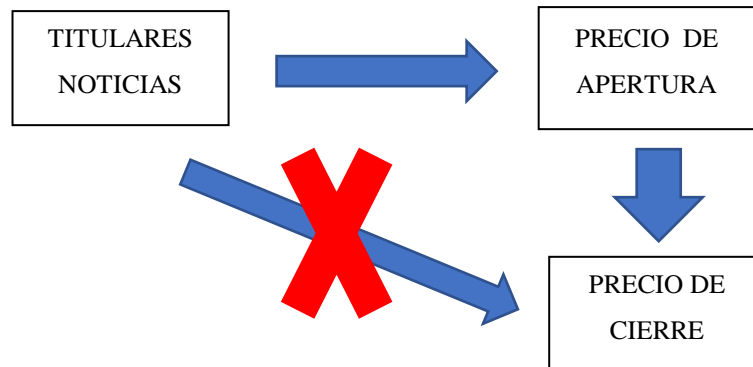
Tabla 7: Modelos precio cierre.

Variable dependiente	l_close			l_close		
	Coeficiente	D. Típica	P-valor	Coeficiente	D. Típica	P-valor
const	0,03348	0,00187	***	0,02933	0,00160	***
pos	0,00054	0,00022	**	0,00022	0,00019	
neg	-0,00032	0,00032		-0,00018	0,00028	
l_high_1	-0,03947	0,00723	***	0,02232	0,00417	***
l_low_1	-0,08349	0,00675	***	-0,04845	0,00432	***
l_volume_1	-0,00026	0,00010	***	-0,00037	0,00009	***
l_adjusted_1	0,00027	0,00072		-0,00049	0,00062	
l_openn_1	0,07166	0,00611	***			
l_close_1	1,04412	0,00643	***	0,07327	0,00589	***
l_openn				0,94764	0,00390	***
N	168.200			168.200		
R ²	0,99959			0,99969		

Fuente: Elaboración propia

Se puede observar en la tabla 7 que ambas variables de interés no son significativas y, por tanto, los titulares de las noticias tanto negativas como positivas tienen un efecto irrelevante en media para el precio de cierre de las empresas del S&P500, además, incluyendo el precio de apertura sin retardo se ha evitado el sesgo. Este resultado se puede deber a que el precio de apertura acumula todo el efecto de las noticias, es decir, si hay una noticia negativa, los inversores en media lo saben antes de que abra el mercado y corrigen sus posiciones antes o a los pocos minutos de la apertura entrando al mercado con una posición corta. En caso de ser una noticia positiva sucedería lo mismo, pero con una posición larga. Además, como las noticias que se emplean son antes de que abra el mercado, esta idea tiene más peso. En definitiva, se concluye que el efecto de los titulares se concentra en el precio de apertura y este sí que afecta directamente al precio de cierre. Si se analizaran noticias que se publican durante la jornada bursátil o al final de ella probablemente sí que afectarían al precio del cierre. En la Figura 2 se puede observar un esquema con la idea que se pretende explicar.

Figura 2: Esquema efecto indirecto de los titulares sobre el cierre



Fuente: Elaboración propia.

En tercer lugar, para explicar el efecto de los titulares sobre la variable de volumen se genera un modelo. Se regresa la variable que contiene el volumen de las transacciones frente a las dos variables binarias que recogen los sentimientos positivos y negativos controlando por el primer retardo de las variables comerciales para explicar la evolución del volumen de las transacciones con el objetivo de eliminar el sesgo por omisión de variable relevante. En la tabla 8 se pueden observar los resultados.

Tabla 8: Modelo volumen de transacciones.

Variable dependiente	l_volume		
	Coeficiente	D. Típica	P-valor
const	6,62072	0,04467	***
pos	0,06490	0,00516	***
neg	0,04919	0,00764	***
l_openn_1	-0,23271	0,14506	
l_high_1	2,34308	0,16965	***
l_low_1	-0,45661	0,16221	***
l_close_1	0,00004	0,00000	***
l_adjusted_1	-1,69398	0,15117	***
l_volume_1	0,55631	0,00238	***
N	168.209		
R ²	0,88123		

Fuente: Elaboración propia

Se puede observar en la tabla 8 que ambas variables de interés son significativas y, por tanto, los titulares de las noticias tanto negativas como positivas tienen un efecto relevante en media para el número de transacciones de las empresas del S&P500. Específicamente si el titular de la noticia es negativo en lugar de ser neutral, el volumen de las transacciones de las empresas del SPY aumentará en media un 4,9%, *ceteris paribus*. Y si el titular de la noticia es positivo en lugar de ser neutral, aumentará en media 6,5%, permaneciendo constante el resto.

Las noticias son pistas de los movimientos de los mercados bursátiles y estas pistas se puede transformar en dinero si se aprovecha la oportunidad. Por este motivo, el número de compraventas aumenta cuando hay noticias porque los inversores intentan aprovecharlas bien invirtiendo con más volumen en la misma cartera de acciones, o invirtiendo en otras empresas por la posibilidad de generar beneficios. Además, la diferencia entre el efecto de los positivos y los negativos se puede deber a que el inversor se siente más seguro y decide invertir más cuando la polaridad del titular es positiva.

En cuarto lugar, para explicar el efecto de los titulares sobre la variable del precio máximo se generan dos nuevos modelos. En el primero, se regresa la variable del precio del precio máximo frente a las dos variables binarias que recogen los sentimientos positivos y negativos controlando por el primer retardo de las variables comerciales para explicar la evolución del precio de apertura con el objetivo de eliminar el sesgo por omisión de variable relevante. En el segundo, se crea el mismo modelo, pero incluyendo la variable del precio de apertura sin retardo con el objetivo de observar si el efecto de los titulares influye más allá de la apertura del mercado y también para evitar el sesgo por omisión de variable relevante. En la tabla 9 se pueden observar los resultados.

Tabla 9: Modelos precio máximo.

Variable dependiente	l_high			l_high		
	Coeficiente	D. Típica	P-valor	Coeficiente	D. Típica	P-valor
const	-0,00283	0,00140	**	-0,00785	0,00104	***
pos	0,00169	0,00016	***	0,00138	0,00012	***
neg	0,00059	0,00024	**	0,00072	0,00018	***
l_low_1	-0,11143	0,00507	***	-0,13714	0,00280	***
l_close_1	0,96770	0,00483	***	0,06113	0,00382	***
l_volume_1	0,00177	0,00008	***	0,00172	0,00006	***
l_adjusted_1	0,00263	0,00054	***	0,00194	0,00040	***
l_openn_1	-0,01096	0,00459	**			
l_openn				0,93381	0,00253	***
l_high_1	0,14834	0,00543	***	0,13776	0,00271	***
N	168.200			168.200		
R ²	0,99977			0,99987		

Fuente: Elaboración propia

Se puede observar en la tabla 9 que ambas variables de interés son significativas y, por tanto, los titulares de las noticias tanto negativas como positivas tienen un efecto relevante en media para el precio máximo de las empresas del índice. Además, se puede observar que el efecto del sentimiento de los titulares tiene efecto después de la apertura ya que, en el segundo modelo, las variables de estudios siguen siendo significativas. También cabe destacar que, al controlar por el precio de apertura sin retardo, el efecto de los titulares positivos disminuye y el de los negativos aumenta, esto puede deberse a que antes los efectos estaban sobre y subestimados, respectivamente. Concretamente, si el titular de la noticia es negativo en lugar de ser neutral, el precio máximo de cotización de las empresas del SPY aumentará en media un 0,07%, *ceteris paribus*. Y si el titular de la noticia es positivo en lugar de ser neutral, aumentará en media 0,14%, permaneciendo constante el resto. Los efectos de ambas variables son positivos, esto puede indicar que cuando hay titulares positivos y negativos, la volatilidad del precio de las acciones se dispara, esta idea queda reforzada porque en el modelo del volumen se ha demostrado que los titulares negativos y positivos aumentan las transacciones del mercado.

En quinto lugar, para explicar el efecto de los titulares sobre la variable del precio mínimo se generan dos nuevos modelos. En el primero, se regresa la variable del precio mínimo frente a las dos variables binarias que recogen los sentimientos positivos y negativos controlando por el primer retardo de las variables comerciales para explicar la evolución del precio de apertura con el objetivo de eliminar el sesgo por omisión de variable relevante. En el segundo, se crea el mismo modelo, pero incluyendo la variable del precio de apertura sin retardo con el objetivo de observar si el efecto de los titulares influye más allá de la apertura del mercado y también para evitar el sesgo por omisión de variable relevante. En la tabla 10 se pueden observar los resultados.

Tabla 10: Modelos precio mínimo.

Variable dependiente	l_low			l_low		
	Coeficiente	D. Típica	P-valor	Coeficiente	D. Típica	P-valor
const	0,03825	0,00154	***	0,03334	0,00116	***
pos	-0,00049	0,00018	***	-0,00082	0,00014	***
neg	-0,00084	0,00027	**	-0,00069	0,00020	***
l_high_1	-0,17819	0,00595	***	-0,15073	0,00301	***
l_close_1	1,06884	0,00530	***	0,06428	0,00424	***
l_volume_1	-0,00172	0,00008	***	-0,00181	0,00006	***
l_adjusted_1	-0,00099	0,00059	*	-0,00177	0,00044	***
l_openn_1	0,03255	0,00503	***			
l_openn				1,00668	0,00281	***
l_low_1	0,07319	0,00556	***	0,07822	0,00311	***
N	168.200			168.200		
R ²	0,99972			0,99984		

Fuente: Elaboración propia

Se puede observar en la tabla 10 que ambas variables de interés son significativas y, por tanto, los titulares de las noticias tanto negativas como positivas tienen un efecto relevante en media para el precio mínimo de las empresas del índice. Además, se puede observar que el efecto del sentimiento de los titulares tiene efecto después de la apertura ya que, en el segundo modelo, las variables de estudios siguen siendo significativas. También cabe destacar que, al controlar por el precio de apertura sin retardo, el efecto de los titulares positivos aumenta y el de los negativos disminuye, esto puede deberse a que

antes los efectos estaban sub y sobreestimados, respectivamente. Concretamente, si el titular de la noticia es negativo en lugar de ser neutral, el precio mínimo de cotización de las empresas del SPY disminuirá en media un 0.07%, *ceteris paribus*. Y si el titular de la noticia es positivo en lugar de ser neutral, disminuirá en media 0,08%, permaneciendo constante el resto. Los efectos de ambas variables son negativos, esto puede indicar la misma idea anteriormente explicada.

A continuación, se van a presentar los resultados de la segunda parte de este estudio empírico. Todos los resultados son filtrados con un nivel de significatividad del 5%.

En primer lugar, se regresa para las 504 empresas la variable del precio de apertura frente a las dos variables binarias que recogen los sentimientos positivos y negativos controlando por el resto de los precios. Se seleccionan los coeficientes de las empresas (β_{pos}) y (β_{neg}) que sean positivos y negativos respectivamente y cumpliendo que uno u otro de los coeficientes sean significativos. Se encuentran 55 empresas que cumplen los requisitos y en la tabla 1 del apéndice aparecen los resultados. Para el precio de cierre se realizan exactamente el mismo procedimiento, se encuentran 18 empresas que cumplen los requisitos y en la tabla 2 del apéndice aparecen los resultados. En el caso del volumen y del precio máximo igual, pero se seleccionan los coeficientes (β_{pos}) y (β_{neg}) que ambos sean positivos. Se encuentran 30 y 78 empresas que cumplen los requisitos y en la tabla 3 y 4 del apéndice aparecen los resultados. Y, por último, en el caso del precio mínimo idéntico, pero se seleccionan los coeficientes que ambos sean positivos, se encuentran 62 empresas que cumplen los requisitos y en la tabla 5 del apéndice aparecen los resultados.

En la tabla 11 se van a presentar los estadísticos principales de los coeficientes (β_{pos}) y (β_{neg}) filtrados de las regresiones anteriormente hechas con el objetivo de observar los datos de una forma compacta y resumida.

Tabla 11: Estadísticos principales efecto individual

Variables	Openn		Close		Volume	
	bpos	bneg	bpos	bneg	bpos	bneg
Mínimo	4,22E-20	-3,46E-14	0	-0,0076	4,62E-18	9,50E-18
1st Qu.	8,99E-19	-6,54E-15	0	-0,0006	4,67E-17	3,80E-17
Mediana	3,04E-18	-2,64E-15	0	0	9,51E-17	7,51E-17
Media	4,32E-18	-5,52E-15	0,0003	-0,0009	1,74E-16	2,61E-16
3rd Qu.	5,93E-18	-1,06E-15	0,0002	0	2,19E-16	2,22E-16
Máximo	1,89E-17	-1,66E-17	0,0020	0	8,36E-16	2,17E-15
N	55		18		30	

Variables	High		Low	
	bpos	bneg	bpos	bneg
Mínimo	5,23E-20	1,95E-19	-1,60E-17	-4,25E-17
1st Qu.	7,14E-19	1,23E-18	-4,23E-18	-5,14E-18
Mediana	2,44E-19	3,73E-18	-1,46E-18	-2,86E-18
Media	4,27E-18	4,94E-18	-2,93E-18	-4,85E-18
3rd Qu.	5,34E-18	7,42E-18	-5,20E-19	-8,09E-19
Máximo	5,98E-17	2,23E-17	-1,45E-20	-2,44E-20
N	78		62	

Fuente: Elaboración propia

Como se pueden observar en la tabla 11, se demuestra una vez más que existen efectos de los titulares de las noticias sobre la evolución de los precios de las cotizaciones de las empresas y que dichos efectos son diferentes para cada empresa. Aunque estos efectos no se dan para todas las empresas del índice y a continuación se va a intentar explicar por qué. Primero se comparan la cantidad de titulares que tienen las empresas que sí que tienen un efecto significativo individual de las que no, con el objetivo de comprobar si las empresas con efecto significativo tienen más titulares con polaridad. En la tabla 12 se muestran los resultados.

Tabla 12: N° de titulares con polaridad

	Openn	Close	Volume	High	Low
Efecto significativo	829	230	624	1.403	1073
Efecto no significativo	6.455	7.054	6.660	5.881	6211

Fuente: Elaboración propia

Como se puede contemplar en la tabla 12, a las empresas que les influye la polaridad del titular de la noticia, tienen menos titulares con contenido positivo y negativo que a las que no les influye. Esto también puede deberse a que las empresas a las que les afecta la polaridad del titular son mucho menos que el total del índice. Otra razón para justificar porque a unas empresas les influyen los titulares y a otras no podría por la intensidad de la polaridad de los titulares, es decir, empresas con titulares más intensos obtendrían efectos. Por este motivo se va a realizar un recuento de las clases de polaridad para los titulares de las empresas con y sin efecto. En la tabla 13 aparecen los resultados. Como se puede comprobar, la hipótesis planteada tampoco se cumple del todo porque las empresas que no tienen efecto tienen siempre mayor número de titulares en todas las

clases de polaridad negativas y positivas. Pero también tienen mayor número de titulares neutrales, lo que puede provocar que el efecto de los titulares positivos y negativos se diluya.

Tabla 13: Clases de polaridad de los titulares

	OPEN		CLOSE		VOLUME		HIGH		LOW	
sent	Con	Sin	Con	Sin	Con	Sin	Con	Sin	Con	Sin
-6	0	2	0	2	1	1	1	1	1	1
-5	4	4	0	8	0	8	2	6	0	8
-4	1	19	1	19	1	19	5	15	6	14
-3	11	61	1	71	6	66	19	53	6	66
-2	34	286	9	311	30	290	81	239	50	270
-1	198	1.626	74	1.750	175	1.649	378	1.446	314	1.510
0	17.650	141.631	5.818	153.463	9.456	149.825	24.804	134.477	467	139.736
1	375	2.909	98	3.183	252	3.029	612	2.669	19.545	2.814
2	140	1.130	37	1.233	121	1.149	223	1.047	151	1.119
3	57	311	9	259	31	337	61	307	54	314
4	9	88	1	96	5	92	14	83	23	74
5	0	17	0	17	1	16	6	11	1	17
6	0	5	0	5	1	4	1	4	0	4

Fuente: Elaboración propia

6. CONCLUSIONES

Internet, las páginas web y las redes sociales hoy en día son una herramienta más para la sociedad. Las empresas también hacen uso de ellas ya que favorecen la toma de decisiones y sus beneficios. Este trabajo se ha centrado en el sector financiero ya que los inversores tienen muy en cuenta las noticias a la hora de tomar sus decisiones bursátiles.

Una de las primeras conclusiones a la que se llega en este trabajo gracias al primer estudio empírico que se realiza es que, mediante la técnica de análisis de sentimiento que proporciona la minería de texto y las herramientas computacionales se ha conseguido extraer la información relevante de una gran base de datos de titulares de noticias, minimizando el tiempo empleado y sintetizando dicha información en una variable.

La segunda conclusión que se obtiene gracias al segundo estudio empírico es que la información recogida mediante el análisis de sentimiento que contienen los titulares de las noticias es relevante para poder explicar la evolución de los precios de las cotizaciones del índice bursátil S&P500. Las conclusiones del estudio del efecto conjunto son las siguientes:

- Los titulares con polaridad positiva aumentan en media el precio de apertura 0,03% en comparación con los neutrales y los negativos no tienen un efecto significativo.
- Ni lo titulares positivos ni negativos tienen un efecto relevante en media sobre el precio del cierre porque el efecto lo recoge el precio de apertura.
- Los titulares positivos y negativos aumentan en media la cantidad de las transacciones un 6,5% y 4,9% en comparación con los titulares neutrales.
- Los titulares positivos y negativos aumentan en media el precio máximo un 0,14% y 0,07% en comparación con los titulares neutrales.
- Los titulares positivos y negativos disminuyen en media el precio mínimo un 0,08% y 0,07% en comparación con los titulares neutrales.

Y las conclusiones del estudio del efecto individual son las siguientes:

- Se ha demostrado que existen efectos de los titulares de las noticias sobre la evolución de los precios de las cotizaciones de las empresas y que dichos efectos son diferentes para cada empresa.

Por este motivo, es importante que los inversores tengan en cuenta las noticias a la hora de crear una estrategia de inversión.

7. BIBLIOGRAFÍA

Bravo-Marquez, F., Mendoza, M. and Poblete, B. (2014) ‘Meta-level sentiment models for big social data analysis’, *Knowledge-Based Systems*. Elsevier, 69, pp. 86–99. doi: 10.1016/J.KNOSYS.2014.05.016.

Chen, Y. and Skiena, S. (2014) ‘Building Sentiment Lexicons for All Major Languages’, *Acl*, pp. 383–389. doi: 10.3115/v1/P14-2063.

Hu, M. and Liu, B. (2004) ‘Mining and summarizing customer reviews’, *Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '04*, p. 168. doi: 10.1145/1014052.1014073.

Im, T. L. *et al.* (2014) ‘Impact of Financial News Headline and Content to Market Sentiment’, *International Journal of Machine Learning and Computing*, 4(3), pp. 237–242. doi: 10.7763/IJMLC.2014.V4.418.

Liu, B. (2012) ‘Sentiment Analysis and Opinion Mining’, *Encyclopedia of Machine Learning and Data Mining*. doi: 10.1007/978-1-4899-7502-7_907-1.

Pang, B., Lee, L. and Vaithyanathan, S. (2002) ‘Thumbs up?: sentiment classification using machine learning techniques’, *Empirical Methods in Natural Language Processing (EMNLP)*, 10(July), pp. 79–86. doi: 10.3115/1118693.1118704.

Strapparava, C. and Mihalcea, R. (2007) ‘Semeval-2007 task 14: Affective text’, *Proc. of SemEval-2007*, (June), pp. 70–74. doi: 10.1145/1363686.1364052.

Turney, P. D. (2002) ‘Thumbs up or thumbs down? Semantic Orientation applied to Unsupervised Classification of Reviews’, *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, (July), pp. 417–424. doi: 10.3115/1073083.1073153.

Valero Moreno, A. I. (2017) ‘Técnicas estadísticas en minería de textos’. Available at: <https://idus.us.es/xmlui/handle/11441/63197>.

Visa, S. *et al.* (2011) ‘Confusion matrix-based feature selection’, *CEUR Workshop Proceedings*, 710, pp. 120–127. doi: 10.1.1.666.8961.

Wooldridge, J. M. (2010) *Introducción a la Econometría 4ta. Edición*.

Gartner IT Glossary. (2018). ‘What Is Big Data? - Gartner IT Glossary - Big Data’. [online] Available at: <https://www.gartner.com/it-glossary/big-data/> [Accessed 21 May 2018].

Pérez López, C. and Santín González, D. (2008). ‘Minería de datos’. [Madrid]: Paraninfo Cengage Learning.

IBM Security (2016). ‘Seguridad cognitiva’.

S. Baumer, Benjamin, T. Kaplan, Daniel, J. Horton, Nicholas. (2017). 'Modern data science with R'

Van De Kauter, M., Breesch, D. and Hoste, V. (2015) 'Fine-grained analysis of explicit and implicit sentiment in financial news articles', *Expert Systems with Applications*. Elsevier Ltd, 42(11), pp. 4999–5010. doi: 10.1016/j.eswa.2015.02.007.

Liu, B. (2016) 'Sentiment Analysis and Opinion Mining', *Encyclopedia of Machine Learning and Data Mining*, (May), pp. 1–10. doi: 10.1007/978-1-4899-7502-7_907-1.

Lubitz, M. (2018) 'Who drives the market? Sentiment analysis of financial news posted on Reddit and the Financial Times'.

Sulaiman, A. (2017) 'Using Tweets Sentiment Analysis to Predict Stock Market Movement'.

Zhang, L. (2013) 'Sentiment Analysis on Twitter with Stock Price and Significant Keyword Correlation', pp. 1–30.

Liu, B. (2018). 'Opinion Mining, Sentiment Analysis, Opinion Extraction'. [online] Cs.uic.edu. Available at: <https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon> [Accessed 10 Apr. 2018].

Martin, B. and Koufos, N. (2018). 'Sentiment Analysis on Reddit News Headlines with Python's Natural Language Toolkit (NLTK) – LearnDataSci'. [online] *Learndatasci.com*. Available at: <https://www.learndatasci.com/tutorials/sentiment-analysis-reddit-headlines-pythons-nltk/#>: [Accessed 16 Apr. 2018].

Sites.google.com. (2018). 'Multilingual Sentiment Analysis Presented at ACL - Data Science Lab'. [online] Available at: <https://sites.google.com/site/datascienceslab/blog-1/multilingualsentimentanalysispresentedatacl> [Accessed 10 Apr. 2018].

Nielsen, F. (2011). AFINN. [online] *Www2.imm.dtu.dk*. Available at: http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=6010 [Accessed 19 Apr. 2018].

Mariapinto.es. (2018). Inquirer Home Page. [online] Available at: <http://www.mariapinto.es/ciberabstracts/Articulos/Inquirer.htm> [Accessed 20 Apr. 2018].

8. APÉNDICE

Tabla 1: Efecto individual sobre el precio de apertura

Openn						
ticker	bpos	ppos	bneg	pneg	nb	n
ABC	1,337E-18	2,15E-05	-9,458E-19	0,031542	19	6
AEE	4,94872E-18	0,000964	-8,9415E-19	0,302862	15	4
AFL	1,96199E-18	1,62E-20	-2,63284E-18	2,69E-12	45	32
AJG	1,20661E-18	0,000104	-8,32782E-19	0,117884	28	8
ANTM	2,01038E-18	0,066168	-6,72315E-18	6,05E-07	26	17
AZO	5,34226E-18	0,030997	-6,48706E-18	0,196243	12	5
CB	2,28892E-19	0,621133	-8,42123E-19	0,029656	20	15
CLX	4,4468E-19	6,7E-16	-8,80252E-19	2,44E-24	37	17
COL	1,09851E-17	3,21E-60	-1,06242E-18	0,050514	18	4
CXO	1,46957E-17	9,94E-05	-1,79115E-18	0,411041	12	8
DIS	4,90101E-19	0,470236	-9,18839E-18	0,0002	17	12
DISCK	4,2163E-20	0,937358	-2,73655E-17	3,4E-100	23	11
EL	1,06051E-17	2E-272	-1,54146E-18	2,58E-37	26	13
FBHS	8,46275E-19	0,185988	-4,20521E-18	0,044386	13	12
FIS	2,25187E-18	0,002439	-2,94858E-18	0,244408	52	28
FLR	5,28734E-19	0,000297	-9,91922E-20	0,594281	31	16
GD	8,1164E-18	7,84E-14	-2,5728E-18	0,252834	10	6
GRMN	8,48571E-20	0,652179	-3,87501E-18	3,33E-08	65	32
HAS	1,33604E-18	7,06E-11	-1,22806E-18	0,000248	28	21
HP	7,23541E-18	0,032556	-7,22135E-18	0,031766	11	2
HRL	2,9994E-18	6,31E-08	-5,28256E-19	0,689573	35	15
IFF	9,07969E-18	0,000638	-2,95422E-18	0,513379	4	4
IPGP	4,71358E-18	0,047871	-8,67693E-18	8,57E-07	17	6
IT	3,31202E-18	1,62E-05	-5,58198E-19	0,617428	29	9
IVZ	5,26492E-19	8,81E-05	-3,08866E-18	2,23E-28	30	19
JBHT	3,80675E-18	0,030407	-6,01579E-18	0,014575	15	9
K	5,46167E-19	0,00723	-7,79605E-19	0,020924	55	32
KR	3,60859E-18	0,001107	-6,59954E-18	0,001309	24	9
LMT	1,29642E-18	0,014821	-1,37675E-19	0,893117	71	30
LNT	2,92058E-18	0,175822	-9,78534E-18	0,039775	12	6
LYB	4,68373E-18	0,007197	-5,09598E-18	0,087368	27	8
MET	1,98539E-19	0,840712	-5,63735E-18	0,004139	44	25
MRK	8,19027E-19	0,075298	-1,89063E-18	0,000602	91	54
NBL	3,35476E-18	4,09E-06	-2,61431E-17	8,96E-17	31	20
NKE	4,69861E-18	9,54E-05	-1,70242E-18	0,306393	3	3
NLSN	1,75977E-18	0,037097	-1,27374E-18	0,381052	45	20
NRG	7,72182E-18	0,064934	-3,46347E-17	0,000117	26	11
NTAP	4,90871E-18	0,002785	-9,86957E-18	0,000392	51	32
PHM	9,52004E-19	7,82E-09	-1,63581E-18	0,001136	26	11
PM	3,90011E-19	0,322263	-1,15684E-18	0,007242	32	22

RRC	1,89449E-17	0,104593	-3,00233E-17	0,000285	7	3
SCG	4,78117E-18	6,19E-08	-3,1012E-18	7,1E-07	34	12
SCHW	1,76186E-18	0,00258	-1,05924E-18	0,217395	37	29
STT	6,51382E-18	0,025348	-2,78342E-18	0,364923	26	15
STX	7,51322E-18	0,031614	-1,26672E-17	0,195368	22	9
T	5,32206E-19	0,0013	-9,76469E-20	0,622295	157	69
TGT	4,93532E-18	0,037587	-1,66228E-20	0,993407	50	19
TMO	3,04362E-18	1,75E-05	-7,40789E-19	0,630318	23	6
TSS	1,22731E-17	1,22E-97	-1,8655E-17	9,31E-47	34	18
VMC	3,6084E-19	0,06047	-1,47448E-18	0,001643	12	7
WAT	1,75744E-17	2,37E-16	-1,8697E-18	0,509234	19	6
WM	8,05949E-18	0,000719	-2,63891E-18	0,001057	20	11
XEL	1,08911E-18	0,54707	-3,13134E-18	0,014414	14	3
ZBH	3,83952E-18	0,039854	-8,81152E-18	0,000155	22	5
ZION	9,38105E-18	0,025247	-5,02907E-18	0,39237	5	3

Fuente: Elaboración propia

Tabla 2: Efecto individual sobre el precio de cierre

close					
	ticker	bpos	ppos	bneg	pneg
1	ADBE	3,28175E-18	2,80545E-05	-2,03445E-18	0,000136926
2	ADI	0,002036386	0,011990425	-0,001063797	0,41016959
3	AEE	4,67019E-18	1,74662E-23	-3,6576E-19	0,147686245
4	ANSS	1,76426E-18	0,02712031	-1,74689E-18	0,356948073
5	AVGO	5,19949E-20	0,983237207	-5,71299E-18	0,002250177
6	AYI	0,00064754	0,248686686	-0,000862156	0,018916626
7	CXO	3,4929E-20	0,955823682	-4,13084E-18	5,52611E-25
8	EA	1,10096E-18	0,001788254	-4,26358E-18	4,73578E-13
9	ETFC	5,47449E-19	0,435064737	-5,96652E-18	3,0861E-17
10	FISV	7,90208E-19	0,001154418	-4,62569E-18	3,97041E-12
11	HBI	0,0004924	0,688301161	-0,007593096	0,009579819
12	HOLX	2,3725E-18	3,12402E-06	-1,37197E-18	0,041027285
13	IVZ	0,001140936	0,021135302	-0,001543271	0,10205016
14	KORS	1,63481E-18	0,131409956	-8,08942E-18	3,57183E-24
15	MTB	0,000222495	0,831799205	-0,006034265	0,000351207
16	PWR	9,28447E-19	0,258527068	-2,48509E-18	0,002648236
17	SIVB	1,19663E-18	9,55098E-05	-1,44835E-18	0,050424757
18	VRSK	4,2245E-18	1,10689E-14	-2,87542E-18	3,95538E-05

Fuente: Elaboración propia

Tabla 3: Efecto individual sobre el volumen

volumen					
	ticker	bpos	ppos	bneg	pneg
1	ADM	8,35866E-16	0,006343148	4,34787E-16	0,35713168
2	ADP	5,41644E-18	0,287797027	1,83588E-17	0,002046105
3	AMGN	6,66888E-18	0,195669825	3,11503E-17	3,86372E-05
4	ARE	2,52315E-16	0,002222821	6,08646E-17	0,549524216
5	BIIB	2,24648E-16	0,018245909	4,72181E-17	0,639204602
6	BLL	1,75411E-16	0,003653244	4,59173E-16	3,88623E-07
7	BMJ	1,13133E-16	0,006660243	1,93228E-17	0,648819085
8	CAH	9,01706E-17	0,000918643	8,76609E-17	0,21454713
9	CELG	7,63472E-17	4,63002E-05	5,55317E-17	0,005218901
10	CRM	6,90139E-17	0,107740615	8,68515E-17	0,045292541
11	CXO	8,84803E-17	0,045767848	1,11672E-16	1,92395E-05
12	EBAY	6,82848E-17	0,001921334	4,44546E-17	0,100999252
13	EMR	4,45722E-17	0,781016125	2,1738E-15	6,27289E-05
14	FOXA	4,24013E-16	5,22038E-08	2,58402E-16	0,013328575
15	GPN	4,8825E-16	0,000343971	1,80848E-16	0,107923901
16	HON	4,39038E-17	0,000428615	9,50011E-18	0,67133696
17	HRL	3,01676E-17	0,022157529	2,4067E-17	0,452763778
18	JEC	1,71202E-17	0,003317616	1,25151E-17	0,244729536
19	KSS	3,45524E-17	0,022793977	1,42627E-16	0,000249025
20	MON	2,90814E-16	0,045729757	1,63088E-15	0,001333007
21	MPC	4,62012E-18	0,662402343	4,62702E-17	0,001757573
22	MSI	1,02862E-16	0,023706645	6,32422E-17	0,328071071
23	NCLH	5,45677E-17	0,247251504	2,08989E-16	0,028014846
24	OMC	5,30788E-17	0,004904649	4,77951E-17	0,239727458
25	SPG	9,99392E-17	0,003287428	3,58621E-17	0,681369007
26	TSN	1,64396E-16	0,014785379	2,26585E-16	0,001842644
27	USB	1,28606E-16	0,417745528	5,02687E-16	0,02544011
28	VNO	3,92933E-16	0,048896473	1,72848E-16	0,534056154
29	WRK	2,00802E-16	2,04554E-05	2,69711E-17	0,67506114
30	XEL	6,43649E-16	0,079003955	6,01717E-16	0,019988782

Fuente: Elaboración propia

Tabla 4: Efecto individual sobre el precio máximo

high					
	ticker	bpos	ppos	bneg	pneg
1	A	7,06E-18	2,6E-06	5,74E-18	9,07E-05
2	ABBV	9,15E-19	0,362601	2,8E-18	0,001228
3	ACN	1,15E-18	0,00029	3,96E-18	4,17E-21
4	ADM	1,03E-17	6,8E-07	8,2E-19	0,794742
5	AEE	1,29E-17	0,006865	5,2E-18	0,061526
6	AIV	7,99E-18	2,7E-23	1,08E-18	0,297706
7	AKAM	1,42E-19	0,956516	9,76E-18	0,009671
8	ALL	5,35E-19	0,115514	1,81E-18	6,24E-06
9	AMP	7,52E-18	3,14E-06	1,08E-17	0,009456
10	AMT	3,62E-18	3,27E-05	1,59E-18	0,234997
11	AON	2,08E-19	0,822423	1,91E-18	0,047353
12	ARNC	4,41E-19	0,852583	1,22E-17	0,022374
13	ATVI	8,41E-18	3,69E-08	4,45E-19	0,871431
14	AWK	2,85E-18	0,123797	7,79E-18	0,007552
15	AYI	5,98E-17	3E-225	7,35E-19	0,123869
16	BIIB	8,18E-19	0,559196	2,93E-18	0,049514
17	CELG	4,66E-18	0,005236	5,74E-18	0,001225
18	CF	1,15E-17	2,5E-107	7,5E-19	0,102941
19	CI	6,37E-18	1,71E-06	9,2E-18	0,000427
20	CTL	6,98E-18	0,002558	2,48E-18	0,494297
21	CVS	6,41E-19	0,007537	3,61E-19	0,308738
22	CXO	1,79E-18	0,000737	4,14E-18	3,08E-33
23	DFS	7,51E-19	0,180204	4,51E-18	5,79E-12
24	DISCA	4,82E-18	0,014896	7,13E-18	0,120524
25	DLR	5,38E-18	0,005011	2,61E-18	0,208743
26	DPS	1,3E-19	0,903262	8,32E-18	1,45E-07
27	DVN	9,56E-19	0,522002	5,82E-18	0,014283
28	DXC	1,51E-17	1,04E-21	1,69E-17	5,13E-26
29	EIX	1,83E-19	0,901645	4,97E-18	0,018567
30	ESRX	8,56E-19	0,151278	6,56E-18	0,00026
31	ETFC	3,76E-18	0,012084	7,51E-18	2,29E-07
32	EXPE	2,84E-18	0,008902	5,84E-19	0,628297
33	FAST	9,36E-18	0,001024	1,2E-17	2,8E-05
34	FRT	3,5E-18	0,005049	1,33E-18	0,449997
35	GPC	2,26E-18	0,000306	2,7E-18	0,203417
36	HIG	1,81E-18	0,006352	6,02E-19	0,423615
37	HOG	6,48E-18	0,002129	1,35E-17	1,12E-17
38	HPE	2,05E-19	0,66866	7,68E-18	5,03E-17
39	HUM	7,76E-18	2,2E-141	1,8E-18	0,002536
40	IRM	2,73E-18	0,000114	1,56E-18	0,165714
41	IVZ	3,97E-19	0,432258	2,49E-18	0,010514
42	JCI	6,09E-18	0,000251	1,88E-18	0,708064

43	JNJ	4,19E-19	0,62371	8,67E-18	7,89E-05
44	K	1,01E-18	0,005624	7,13E-19	0,239397
45	KHC	5,2E-18	2,13E-05	9,41E-18	6,26E-06
46	KORS	3,08E-18	0,002035	3,87E-18	1,49E-08
47	LUV	5,85E-18	1,59E-05	5,7E-18	0,00266
48	MAR	1,89E-19	0,82344	1,4E-17	2,4E-20
49	MET	5,52E-19	0,481494	4,38E-18	0,005084
50	MRK	7,02E-19	6,34E-06	6,37E-19	0,000532
51	MYL	7,61E-18	7,38E-07	1,56E-18	0,292673
52	NEE	3,57E-18	5,07E-16	1,01E-18	0,024708
53	NKTR	1,83E-17	1,98E-86	3,86E-18	8,24E-13
54	NLSN	1,38E-18	0,00684	1,54E-18	0,078935
55	NRG	4,49E-18	0,029265	2,23E-17	5,35E-07
56	PH	3,79E-18	4,61E-05	1,2E-18	0,19108
57	PKI	2,47E-18	0,104165	1,34E-17	0,000926
58	PNC	3,87E-18	3,82E-05	3,59E-18	0,00627
59	RF	5,63E-19	0,533932	2,77E-18	0,027896
60	RHT	5,23E-20	0,787316	8,9E-19	0,03723
61	SEE	6,97E-18	3,73E-14	5,13E-19	0,767645
62	STT	7,92E-19	0,355217	4,25E-18	3,76E-06
63	T	1,73E-18	5,87E-09	3,62E-19	0,299732
64	TAP	1,41E-19	0,961322	9,01E-18	0,002262
65	TDG	5,31E-19	0,851436	6,11E-18	0,002574
66	TGT	1,96E-18	0,001707	1,95E-19	0,712505
67	TROW	4,3E-18	0,000234	1,51E-18	0,446225
68	TXT	8,4E-19	0,033852	5,36E-18	6,34E-17
69	V	1,49E-18	0,013947	6,98E-18	2,97E-13
70	VAR	3,84E-18	3,31E-06	7,71E-19	0,48464
71	WAT	4,48E-18	5,19E-08	4,61E-19	0,677318
72	WBA	7,59E-19	0,375757	1,1E-17	1,07E-14
73	WFC	2,18E-19	0,453218	2,2E-18	7,81E-05
74	WY	3,44E-19	0,816628	5,37E-18	0,000292
75	WYNN	4,05E-19	0,917778	1,87E-17	3,31E-08
76	XEC	1,7E-18	0,020753	5,57E-19	0,449433
77	XYL	2,41E-18	0,00031	6,66E-19	0,556023
78	ZION	5,18E-18	0,071933	9,12E-18	0,024444

Fuente: Elaboración propia

Tabla 5: Efecto individual sobre el precio mínimo

low					
	ticker	bpos	ppos	bneg	pneg
1	ADS	-2,51E-19	0,4537832	-7,18E-18	7,92E-18
2	AIZ	-2,89E-18	7,002E-25	-1,45E-18	1,48E-06
3	ALK	-7,99E-19	0,5676982	-6,76E-18	0,000606
4	ALLE	-1,34E-18	0,0323383	-2,27E-18	0,010093
5	ARE	-7,32E-19	0,3102163	-1,89E-18	0,03568
6	AVY	-3,17E-18	0,0022355	-4,6E-19	0,655891
7	BDX	-9,13E-19	0,0147911	-2,91E-18	8,65E-07
8	BKNG	-8,05E-18	1,142E-57	-4,54E-18	6,56E-16
9	CA	-5,29E-19	1,849E-09	-6,4E-19	6,26E-10
10	CBOE	-6,52E-18	6,734E-05	-4,25E-17	1,03E-40
11	CELG	-1,41E-18	1,449E-06	-3,74E-18	2,05E-28
12	CRM	-3,59E-18	0,0092644	-1,22E-18	0,379685
13	DISCA	-4,32E-18	8,307E-05	-5,86E-18	0,020461
14	DISH	-1,42E-19	0,9094083	-4,07E-18	0,007876
15	DPS	-1,54E-18	0,0854778	-1,84E-17	5,22E-36
16	DVN	-7,55E-19	3,105E-08	-1,4E-18	1,15E-10
17	EFX	-1,49E-18	8,054E-12	-4,22E-19	0,022717
18	EIX	-2,37E-19	0,8419659	-4,98E-18	0,0034
19	EL	-4,74E-19	0,2066513	-4,22E-18	9,05E-19
20	FFIV	-1,95E-18	0,3975706	-5,19E-18	0,024307
21	FIS	-8,15E-19	0,0048586	-3,11E-19	0,752459
22	FOX	-4,56E-18	0,0190499	-2,41E-18	0,460135
23	FRT	-3,74E-18	2,336E-09	-7,86E-19	0,361093
24	GM	-5,03E-19	0,1234358	-8,53E-18	2,92E-45
25	GPS	-9,52E-18	6,033E-68	-2,8E-18	1,54E-10
26	HIG	-2,78E-18	1,241E-14	-5,03E-19	0,199123
27	HLT	-3,37E-19	0,6976555	-3,32E-18	0,001345
28	HOG	-6,84E-18	0,0007926	-1,41E-17	5,23E-20
29	HP	-9,34E-18	3,621E-09	-8,69E-18	2,97E-08
30	HPQ	-4,22E-18	0,0004035	-1,37E-17	7,88E-11
31	HRS	-9,27E-19	0,0189532	-2,14E-19	0,834372
32	HST	-7,8E-18	0,0014769	-5,87E-18	0,061131
33	ICE	-1,22E-19	0,8906845	-3,22E-18	0,002371
34	JBHT	-5,15E-18	2,186E-05	-9,18E-18	7,95E-08
35	JPM	-2,61E-19	0,0493541	-3,32E-19	0,087913
36	KORS	-4,29E-18	5,367E-06	-5,04E-19	0,423739
37	KR	-1,43E-18	0,0429804	-3,08E-18	0,019179
38	LEN	-5,84E-18	1,621E-23	-9,95E-19	0,188437
39	LOW	-6,95E-20	0,9608187	-1,87E-17	0,001105
40	LUK	-1,45E-20	0,9609417	-2,35E-18	0,007657
41	MAR	-2,68E-18	3,374E-08	-2,07E-17	7,62E-82
42	MET	-3,84E-19	0,5207815	-3,51E-18	0,003222

43	MPC	-5,27E-18	2,076E-14	-1,3E-17	3,59E-36
44	NDAQ	-2,92E-18	0,013931	-4,36E-20	0,98004
45	NKE	-1,31E-17	8,348E-06	-1,34E-18	0,739927
46	PLD	-1,74E-19	0,8338802	-1,71E-18	0,008221
47	PVH	-1,22E-18	0,1478885	-4,04E-18	0,000707
48	QCOM	-1,29E-18	2,864E-05	-7,58E-19	0,091437
49	QRVO	-9,98E-19	0,0193849	-1,15E-18	0,354087
50	REG	-6,45E-19	0,5082504	-4,81E-18	0,006648
51	RF	-6,74E-20	0,9513311	-3,1E-18	0,043894
52	RRC	-1,6E-17	0,0213815	-1,19E-17	0,014863
53	SLG	-2,36E-18	0,0128178	-5,95E-19	0,100216
54	SPGI	-2,44E-18	0,003915	-3,82E-18	6,96E-05
55	STI	-1,76E-18	0,0443205	-5,8E-19	0,761157
56	T	-5,17E-19	0,0001033	-9,48E-20	0,550527
57	TMO	-2,85E-18	0,043684	-8,77E-19	0,777396
58	TSCO	-4,23E-18	0,0186381	-3,76E-18	0,000997
59	TSN	-7,63E-19	0,010098	-7,16E-19	0,024633
60	VZ	-3,69E-19	0,5034051	-1,92E-18	0,026534
61	WFC	-3,6E-19	0,4675354	-2,45E-18	0,009411
62	ZTS	-1,15E-17	4,353E-11	-2,44E-20	0,977014

Fuente: Elaboración propia